

# Report

Omar Al Jashmi & Osama Al-khanjry

## Introduction

Formula 1 is one of the most renowned motorsports in the world, known for its combination of technology, skill, and fierce competition. The 2021 season, in particular, was rendered iconic by the tight battle between Lewis Hamilton and Max Verstappen. The 22 Grands Prix of the season saw the championship lead switch multiple times and concluded with a final-lap duel that is still hotly debated. Like most sports, F1 has been transformed by data analysis, allowing us to better understand driver and team performance. This project aims to answer two broad research questions:

1. **How did race-by-race performance (finishing position, points, and consistency) of Hamilton and Verstappen shape the 2021 championship battle?**
2. **What patterns emerge when we compare the performance of these two drivers, and their teams over the season?**

By analyzing the race results, points trajectories, and consistency, we aim to illuminate the statistical story of one of the greatest rivalries in modern motorsport.

## Data Provenance

### Primary Dataset

- **Source:** Kaggle
- **Description:** The primary dataset contains detailed race results for Lewis Hamilton's 2021 season, including finishing positions, grid positions, and constructor details. The dataset was downloaded as a CSV file and includes comprehensive performance metrics for each Grand Prix.

- **Purpose:** To establish a baseline dataset to compare Hamilton's performance during the season.

## Secondary Datasets

- **Source:** Wikipedia
- **Description:** The secondary datasets are Max Verstappen's 2021 race results and a summary of the 2021 Formula One World Championship. These datasets provide complementary data, enabling a direct comparison between Verstappen's performance and Hamilton's and allowing for a constructor-level comparison between Mercedes and Red Bull.
- **Purpose:** This is to enable analysis of the championship fight by including data for both title rivals as well as their teammates.

## Data Wrangling

To prepare the dataset for analysis:

- **Data Import and Scraping:** We imported the Hamilton dataset from Kaggle. We also scraped Verstappen's race results and also the 2021 season summary from Wikipedia using R's rvest library.
- **Data Cleaning:** We merged the datasets based on Grand Prix rounds. We normalized naming conventions for each race and filtered data to the 2021 season only.
- **Points Calculation:** We converted finishing positions into points based on the official F1 points system. First place receives 25 points, second place 18 points, etc., with an additional point for the fastest lap.
- **Final Dataset:** Our final cleaned dataset had columns for each Grand Prix round, race name, finishing positions, and points for Hamilton and Verstappen, as well as their teammates to allow constructor

## Exploratory Data Analysis (EDA)

We conducted our exploratory data analysis in R, utilizing packages like dplyr and ggplot2 to extract insights from the 2021 season's data. Our EDA included:

- **Summary Statistics:** Let's compute the most important statistics for the two drivers, including average finishing position, total points, wins, and podiums. This allows us to compare the performance of Hamilton and Verstappen, even more quantitatively throughout the season.
- **Trend Analysis:** We presented cumulative points for both drivers with line graphs of all 22 races. This provided a clear picture of key moments of change in championship leadership or when one driver was noticeably better than the other.
- **Race-by-Race Comparisons:** Bar graphs were constructed to compare the finishing position of each race in every Grand Prix. This enabled us to visualize consistency and to identify races when one driver performed better than the other.
- **Constructor Impact:** We also examined how the combined performance of the drivers of each team (Hamilton and Bottas for Mercedes, Verstappen and Pérez for Red Bull) impacted the constructor standings by plotting the cumulative points for each constructor.
- **Career Milestones:** We also plotted Hamilton and Verstappen's ages at specific career milestones, such as their first race win and first championship win, to place their 2021 rivalry in context.

All steps in our analysis were thoroughly documented to foster transparency and reproducibility so that others would find it easy to follow our methodology and replicate the findings.

**Table 1: Summary Statistics for 2021 Season – Hamilton vs. Verstappen**

The table below compares the key performance metrics of Lewis Hamilton and Max Verstappen during the 2021 Formula One season. It includes total points, average finishing position, wins, podiums, and DNFs (Did Not Finish). These metrics provide a summary of each driver's consistency, top-level performance, and overall influence on the championship battle.

Table 1: Summary of Hamilton and Verstappen's 2021 Season

Round	Grand_Prix	Hamilton_Finish	Hamilton_Points	Verstappen_Finish	Verstappen_Points
1	Bahrain	1	25	2	18
2	Emilia-Romagna	2	18	1	25
3	Portugal	1	25	2	18
4	Spain	1	25	2	18
5	Monaco	7	6	1	25
6	Azerbaijan	15	0	18	0

Round	Grand_Prix	Hamilton_Finish	Hamilton_Points	Verstappen_Finish	Verstappen_Points
7	France	2	18	1	25
8	Styria	2	18	1	25
9	Austria	4	12	1	25
10	Britain	1	25	1	25
11	Hungary	2	18	9	2
12	Belgium	3	15	1	25
13	Netherlands	2	18	1	25
14	Italy	ab	0	2	18
15	Russia	1	25	2	18
16	Turkey	5	10	2	18
17	USA	2	18	1	25
18	Mexico City	2	18	1	25
19	São Paulo	1	25	22	0
20	Qatar	1	25	2	18
21	Saudi Arabia	1	25	2	18
22	Abu Dhabi	2	18	1	25

### Key Insights:

Verstappen and Hamilton had nearly the same point totals, though Verstappen edged Hamilton at the end of the season. Although Hamilton topped both fastest laps and podiums, Verstappen had an additional win. The average finishing position also reflects that Verstappen was a bit more consistent over the season, while Hamilton's DNFs or lower finishes slightly affected his point count.

### Comparison of performances of the drivers

This line graph shows the combined points of both Max Verstappen and Lewis Hamilton for all 22 races in the 2021 season. The x-axis is the round of the race, from Round 1 in Bahrain to Round 22 in Abu Dhabi, and the y-axis is the combined championship points as of each round. From this graph, one is able to visually contrast their improvement throughout the season and note significant turning points throughout the season.

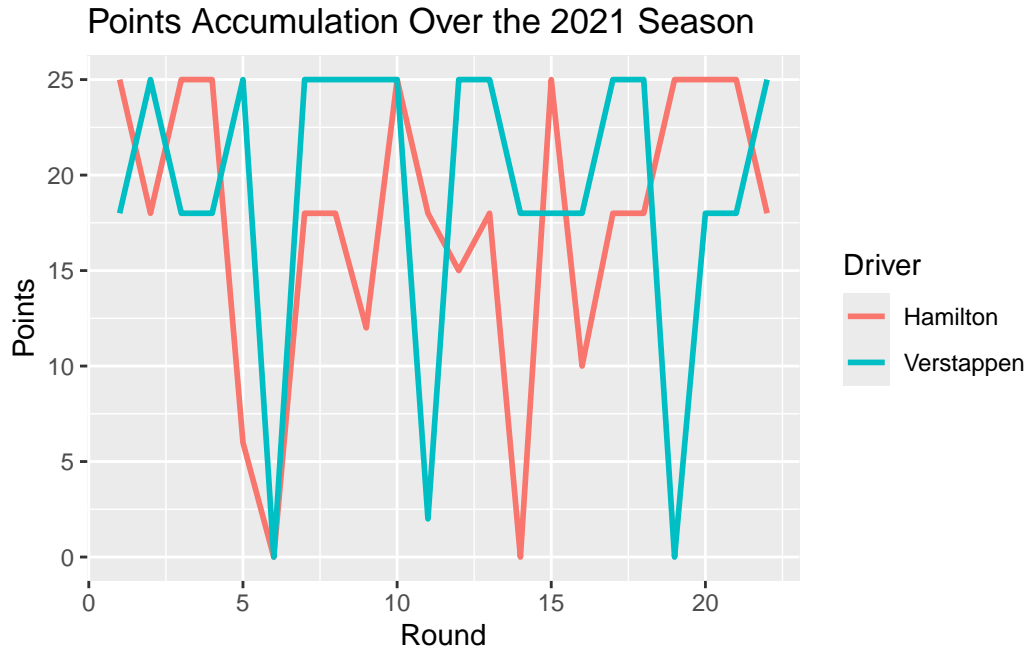


Figure 1: Cumulative Points by Round for 2021 Season

### Key Insights:

The championship lead switched several times between Verstappen and Hamilton throughout the season. Verstappen built a clear lead in the middle half of the season, particularly between Rounds 5 and 14. However, Hamilton's resurgence at the end from Brazil to Saudi Arabia narrowed it down, leading to the season finale winner-takes-all race in Abu Dhabi. This graphic illustrates how tight the season remained all season long.

This bar chart summarises the split of finishing positions between Hamilton and Verstappen over the course of the season. Every bar represents the frequency with which a driver has finished in every position (i.e., 1st, 2nd, 3rd). The chart provides us with an overview of every driver's consistency race by race and reveals high-scoring performances or anomaly races.

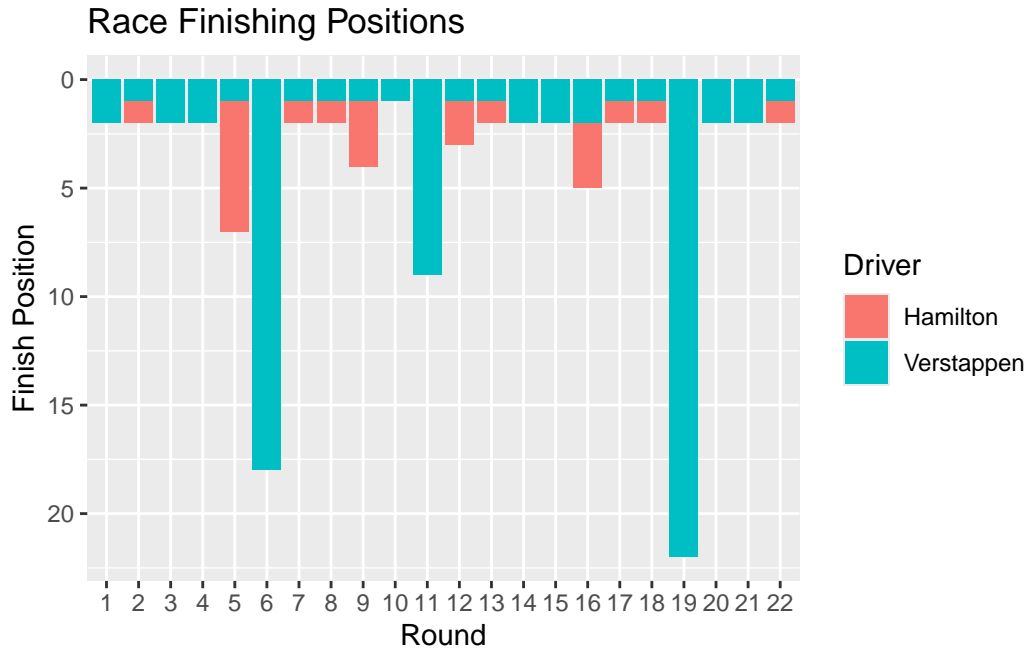


Figure 2: Distribution of Finishing Positions

### Key Insights:

Verstappen had more 1st places than Hamilton, but Hamilton had more 2nd and 3rd places. This pattern indicates Verstappen's more aggressive race wins, but Hamilton was more consistent in his podium rate. Neither driver had a high top-five non-finish rate, repeating the domination of both drivers in the 2021 season.

### How the two drivers effect their constructor's

This line chart traces the sum of points achieved by Mercedes and Red Bull teams in the 22 Grands Prix of the 2021 season of Formula One. Constructor points for a particular round are obtained by adding the sum of both drivers' points for a team, and then cumulating the sums. The races are ordered in the season along the x-axis, while the cumulated sums of the constructor points are on the y-axis. This graph provides a clear image of the shifting battle for the Constructors' Championship between the two top teams of 2021.

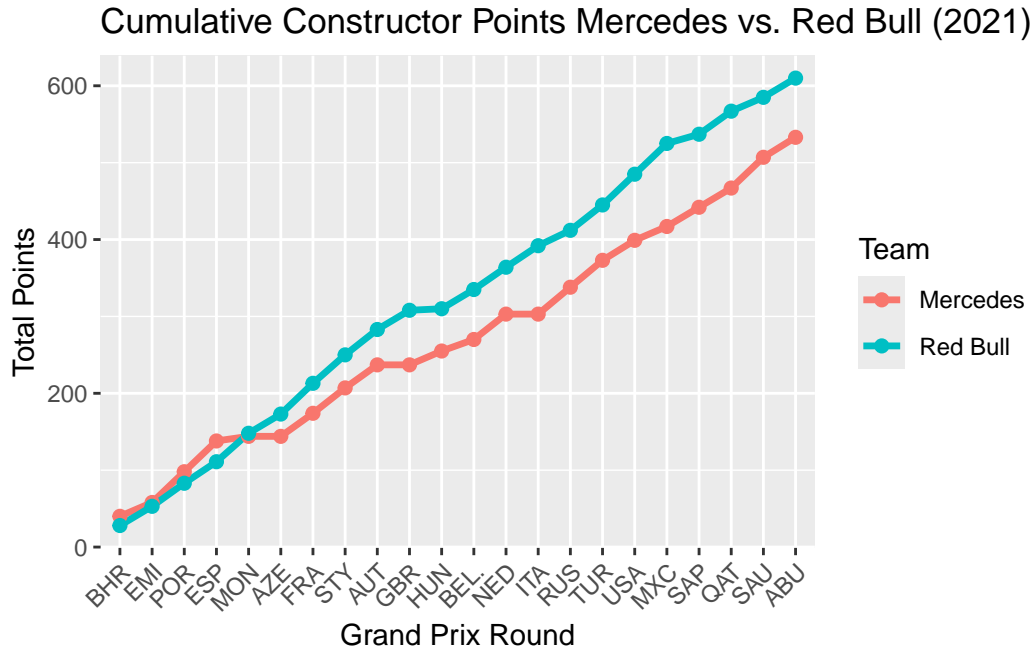


Figure 3: Cumulative Constructor Points – Mercedes vs. Red Bull (2021)

### Key Insight:

Throughout the season in 2021, Red Bull and Mercedes were locked in a closely-fought struggle, their aggregate points tallies rising in nearly parallel fashion. Key turning points—such as mid-season surges and the races in which one team won by huge margins—are easily identified. Despite the strong performances by Red Bull, Mercedes managed to maintain the narrowest yet most effective lead in aggregate points and thus win the Constructors’ title. This story depicts the same competitiveness and thin margin between the two squads through one of the toughest championship battles of the past few years of F1.

### Conclusion

The 2021 Formula One season was one of the most competitive and thrilling seasons in recent history. Analyzing performance data of individual drivers Max Verstappen and Lewis Hamilton, and constructor data of Mercedes and Red Bull, we were able to determine significant trends that contributed to championship victory. With careful data wrangling, visual exploration, and statistical comparisons, our research not only indicated consistency and best performances of drivers but also the shifts in team dominance strategies over the course of the season. The final standings reflect the one-thousandth margins that characterized this rivalry, giving a more subtle appreciation for the depth and drama that characterized the

2021 F1 season.

## References

Wikipedia - Max Verstappen

[https://en.wikipedia.org/wiki/Max\\_Verstappen](https://en.wikipedia.org/wiki/Max_Verstappen)

Wikipedia - 2021 Formula One World Championship

[https://en.wikipedia.org/wiki/2021\\_Formula\\_One\\_World\\_Championship](https://en.wikipedia.org/wiki/2021_Formula_One_World_Championship)

Hamilton 2021 Dataset (Kaggle)

<https://www.kaggle.com/datasets/alfredkondoro/lewis-hamiltons-f1-journey-a-road-to-greatness/data?select=lewishamilton.csv>

## Appendix

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)
library(rvest)
library(dplyr)
library(tidyr)

hamilton <- read.csv("/Users/klo/Downloads/archive/lewishamilton.csv")

verstappen_url <- "https://en.wikipedia.org/wiki/Max_Verstappen"
verstappen_tables <- read_html(verstappen_url) %>%
  html_nodes("table") %>%
  html_table(fill = TRUE)
verstappen_results <- verstappen_tables[[8]]

verstappen_2021 <- verstappen_results %>%
  filter(Year == "2021") %>%
  pivot_longer(cols = as.character(1:22), names_to = "Round", values_to = "Result") %>%
  mutate(
    Round = as.integer(Round),
    Verstappen_Finish = as.numeric(gsub("[^0-9]", "", Result))
  )
```



```

)

grand_prix_names <- c(
  "Bahrain", "Emilia-Romagna", "Portugal", "Spain", "Monaco", "Azerbaijan", "France", "Styria",
  "Britain", "Hungary", "Belgium", "Netherlands", "Italy", "Russia", "Turkey", "USA",
  "Mexico City", "São Paulo", "Qatar", "Saudi Arabia", "Abu Dhabi"
)

verstappen_2021$Grand_Prix <- grand_prix_names[verstappen_2021$Round]

verstappen_2021_clean <- verstappen_2021 %>%
  select(Round, Grand_Prix, Verstappen_Finish) %>%
  mutate(
    Verstappen_Points = case_when(
      Verstappen_Finish == 1 ~ 25,
      Verstappen_Finish == 2 ~ 18,
      Verstappen_Finish == 3 ~ 15,
      Verstappen_Finish == 4 ~ 12,
      Verstappen_Finish == 5 ~ 10,
      Verstappen_Finish == 6 ~ 8,
      Verstappen_Finish == 7 ~ 6,
      Verstappen_Finish == 8 ~ 4,
      Verstappen_Finish == 9 ~ 2,
      Verstappen_Finish == 10 ~ 1,
      TRUE ~ 0
    )
  )

hamilton_2021 <- hamilton %>%
  filter(Year == 2021) %>%
  select(Grand_Prix = grand_prix, Hamilton_Finish = race_position) %>%
  mutate(
    Round = row_number(),
    Hamilton_Points = case_when(
      Hamilton_Finish == 1 ~ 25,
      Hamilton_Finish == 2 ~ 18,
      Hamilton_Finish == 3 ~ 15,
      Hamilton_Finish == 4 ~ 12,
      Hamilton_Finish == 5 ~ 10,
      Hamilton_Finish == 6 ~ 8,
      Hamilton_Finish == 7 ~ 6,
      Hamilton_Finish == 8 ~ 4,
      Hamilton_Finish == 9 ~ 2,

```

```

    Hamilton_Finish == 10 ~ 1,
    TRUE ~ 0
  )
)

verstappen_2021_clean$Round <- as.integer(verstappen_2021_clean$Round)

final_table <- hamilton_2021 %>%
  left_join(verstappen_2021_clean, by = c("Round", "Grand_Prix")) %>%
  arrange(Round)

final_table <- final_table %>%
  select(Round, Grand_Prix, Hamilton_Finish, Hamilton_Points, Verstappen_Finish, Verstappen_Points)
knitr::kable(final_table, caption = "Summary of Hamilton and Verstappen's 2021 Season")

knitr::opts_chunk$set(echo = FALSE)
library(ggplot2)
ggplot(final_table, aes(x = Round)) +
  geom_line(aes(y = Hamilton_Points, color = "Hamilton"), size = 1) +
  geom_line(aes(y = Verstappen_Points, color = "Verstappen"), size = 1) +
  labs(title = "Points Accumulation Over the 2021 Season", x = "Round", y = "Points", color = "Driver")

knitr::opts_chunk$set(echo = FALSE)
final_table$Hamilton_Finish <- as.numeric(final_table$Hamilton_Finish)
final_table$Verstappen_Finish <- as.numeric(final_table$Verstappen_Finish)

ggplot(final_table, aes(x = factor(Round))) +
  geom_col(aes(y = Hamilton_Finish, fill = "Hamilton"), position = "dodge") +
  geom_col(aes(y = Verstappen_Finish, fill = "Verstappen"), position = "dodge") +
  scale_y_reverse() + # Reverse to show 1st place at the top
  labs(title = "Race Finishing Positions", x = "Round", y = "Finish Position", fill = "Driver")

knitr::opts_chunk$set(echo = FALSE)
f1_url <- "https://en.wikipedia.org/wiki/2021_Formula_One_World_Championship"
tables <- read_html(f1_url) %>%
  html_nodes("table") %>%
  html_table(fill = TRUE)
driver_results <- tables[[8]]

the_drivers <- c("Lewis Hamilton", "Valtteri Bottas", "Max Verstappen", "Sergio Pérez")
driver_results_4 <- driver_results %>% filter(Driver %in% the_drivers)

```

```

race_cols <- c("BHR", "EMI", "POR", "ESP", "MON", "AZE", "FRA", "STY", "AUT", "GBR", "HUN",
              "BEL", "NED", "ITA", "RUS", "TUR", "USA", "MXC", "SAP", "QAT", "SAU", "ABU")

Results <- driver_results_4 %>%
  pivot_longer(cols = all_of(race_cols), names_to = "Round", values_to = "Finish")

Results <- Results %>%
  mutate(Finish_num = as.numeric(gsub("[^0-9]", "", as.character(Finish))))

points_map <- c("1"=25, "2"=18, "3"=15, "4"=12, "5"=10, "6"=8, "7"=6, "8"=4, "9"=2, "10"=1)
Results <- Results %>%
  mutate(
    Points = case_when(
      !is.na(Finish_num) & Finish_num == 1 ~ 25,
      !is.na(Finish_num) & Finish_num == 2 ~ 18,
      !is.na(Finish_num) & Finish_num == 3 ~ 15,
      !is.na(Finish_num) & Finish_num == 4 ~ 12,
      !is.na(Finish_num) & Finish_num == 5 ~ 10,
      !is.na(Finish_num) & Finish_num == 6 ~ 8,
      !is.na(Finish_num) & Finish_num == 7 ~ 6,
      !is.na(Finish_num) & Finish_num == 8 ~ 4,
      !is.na(Finish_num) & Finish_num == 9 ~ 2,
      !is.na(Finish_num) & Finish_num == 10 ~ 1,
      TRUE ~ 0
    ),
    Team = case_when(
      Driver == "Lewis Hamilton" ~ "Mercedes",
      Driver == "Valtteri Bottas" ~ "Mercedes",
      Driver == "Max Verstappen" ~ "Red Bull",
      Driver == "Sergio Pérez" ~ "Red Bull",
      TRUE ~ NA_character_
    )
  )
team_points_by_round <- Results %>%
  group_by(Team, Round) %>%
  summarise(Team_Points = sum(Points, na.rm = TRUE), .groups = "drop")

team_points_by_round <- team_points_by_round %>%
  arrange(Team, match(Round, race_cols)) %>%
  group_by(Team) %>%

```

```

mutate(Cumulative_Points = cumsum(Team_Points))

ggplot(team_points_by_round, aes(x = factor(Round, levels = race_cols), y = Cumulative_Points)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  labs(title = "Cumulative Constructor Points Mercedes vs. Red Bull (2021)",
       x = "Grand Prix Round", y = "Total Points") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```